

Figure 3. Overview of the HSV-1 genome depicting coding regions, noncoding features, polymorphisms, and SSRs.

(A) Locations of UL, US, IRL, and IRS in the genome of HSV-1 reference strain 17 (TRL and TRS are omitted). (B) Graph plotting the number of DNA polymorphisms per 500 bp (non-gapped columns) in a whole-genome alignment of 26 HSV-1 sequences (from Table 1). (C) Well-known features of the HSV-1 reference genome are shown mapped to the two DNA strands. These include ORFs, the latency-associated transcript (LAT), untranslated regions (UTRs), origins of DNA replication (OriL and two copies of OriS), and miRNAs (mirs). Widely recognized protein names (e.g., gB encoded by UL27) are included. (D) Location of SSRs plotted along the reference genome, with homopolymers (the same nucleotide repeated ≥ 6 times in a row) plotted separately from larger microsatellites (repeating unit of 2-9 bp) and minisatellites (repeating unit of ≥ 10 bp). SSRs are color-coded to distinguish those for which length is conserved in at least half of the 26 strains (green), vs. those for which length is variable in a majority of strains (orange). Gray SSRs (marked by gray arrowheads) were not coded for conservation, since their length could not be determined by high-throughput sequencing in a majority of strains.